The new participant role scales: A comparison between various criteria for assigning roles and indications for their validity

Goossens, F.A.; Olthof, T.; Dekker, M.C.

published in
Aggressive Behavior
2006

DOI (link to publisher)
10.1002/ab.20133

document version
Publisher's PDF, also known as Version of record

AGGRESSIVE BEHAVIOR
Volume 32, pages 343–357 (2006)

New Participant Role Scales: Comparison Between Various Criteria for Assigning Roles and Indications for Their Validity

F.A. Goossens[1,2]*, T. Olthof[2], and P.H. Dekker[3]

[1]*Department of Special Education, Free University of Amsterdam, Amsterdam, The Netherlands*
[2]*Department of Developmental Psychology, Free University of Amsterdam, Amsterdam, The Netherlands*
[3]*Department of Work and Organizational Psychology, Free University of Amsterdam, Amsterdam, The Netherlands*

The present study was set up with two aims in mind. First, to assess the psychometric characteristics of a peer-report measure of bullying in a Dutch sample, and second, to compare relative and absolute ways of assigning roles in the bullying process. The sample consisted of 242 children (51% boys; mean age approximately 10 years) at T1. Two years later, there were 247 children (49% boys).

We made use of an adaptation of the original Participant Role Scales (PRS) Salmivalli, Lagerspetz, Björkqvist et al., 1996; Salmivalli, Lappalainen and Lagerspetz, 1998) and of the Aggression and Victimization Scale [Perry, Kusel and Perry, 1988]. This scale, called the New PRS, consisted of 32 items in total. Five highly reliable scales were distinguished with the help of CFA, one for leader-like bullying behavior, one for follower-like bullying behavior, and scales for outsider, defender and victim. We computed the roles according to four criteria, that is, *z*-scores (a relative measure) and three different percentage scores (10%, 15% and 20%; each an absolute measure). Sociometric status was also assessed. The concordance between the various methods was moderate. Test–retest stability was also moderate. Test–retest coefficients for the scale scores were considerably higher. Links were found between the roles and sociometric status, irrespective of the method used for assigning roles. Victims were rejected, as were bullies, but defenders were popular. Gender differences indicated that boys were more often bullies or followers, and girls were more often outsiders or defenders. Absolute methods for assigning bullying roles produced fewer unclassifiable children and brought to light substantial differences between classes. Aggr. Behav. 32:343–357, 2006. © **2006 Wiley-Liss, Inc.**

Keywords: bullying; victimization; stability of bullying roles; sociometric status

INTRODUCTION

Interest in the study of bullying has in the past 10 years become truly international, with contributions from the United Kingdom [Smith and Sharp, 1994; Sutton et al., 1999; Wolke et al., 2000], continental Europe [Alsaker and Valkanover, 2001; Björkqvist et al., 1992; Camodeca et al., 2002; Menesini et al., 1997; Olweus, 2001], the United States [Crick and Bigbee, 1998; Juvonen and Graham, 2001; Kochenderfer and Ladd, 1997; Pellegrini, 2001, 2002; Schwartz and Proctor, 2001; Schwartz et al., 2001a,b], Canada [Pepler et al., 1998], Asia [Morita et al., 1999; Schwartz et al., 2001a,b] and Australia [Rigby, 1996; Slee, 1995].

One issue that remains important is how to measure involvement in bullying. In this paper we argue for the use of peers as informants and for the use of absolute criteria (percentage scores) rather than relative criteria (*z*-scores).

Several authors have made use of a self-report measure, developed by Olweus [1989], to uncover involvement in bullying as either a bully or a victim [Ahmad and Smith, 1994; Menesini et al., 1997; Olweus, 1991; Whitney and Smith, 1993]. These authors have frequently argued for the use of anonymous self-reports as they are accurate about the prevalence of bullying, and about other characteristics such as where the bullying takes place and how children respond to being bullied. Crick and Bigbee [1998] mention three other advantages of self-report measures: (1) they can reveal victimiza-

*Correspondence to: F.A. Goossens, Department of Special Education, Free University of Amsterdam, an der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. E-mail: fa.goossens@psy.vu.nl

Received 9 February 2004; Accepted 16 September 2004

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ab.20133

tion episodes no others (except the bully) have been witness to; (2) they can be used in clinical settings where peer reports are unavailable, and (3) they are less time-consuming and easier to administer than peer reports. Nevertheless, there are certain disadvantages to self-report measures. Both Perry et al. [1988] and Juvonen et al. [2000] claim that some children are more likely to depict themselves in a favorable light and may refrain from disclosing their active involvement as a bully or exaggerate their involvement as a victim. For instance, Juvonen et al. showed that using a criterion of .5 SD above the mean for inclusion, about 14% of children could be identified as "true" victims when both self-reports and peer reports were used. These true victims were victims both according to their peers and according to themselves. However, an additional 23% of the subjects saw themselves as victims but were not identified as such by their peers. In contrast with these "paranoids", there were also 7% peer-identified but not self-identified victims ("deniers"). Sutton and Smith [1999] found that in 70% of the cases in their sample the peer reports differed from the self-reports, and the self-reports generally presented a more attractive picture of the roles played in bullying than the peer reports did. It is possible that while many children experience a form of peer harassment, their plight is not always visible to all others. On the other hand, we cannot rule out the possibility that children tend to depict themselves in a favorable light when asked about their involvement in bullying. One argument for the use of peer reports was provided by Pellegrini and Bartini [2000; Pellegrini, 2001], who compared self-reports, peer reports, systematic observations and diaries. The peer-reports method had on average the highest correlations with the other three methods, and was the only method that was significantly correlated with all other methods. Moreover, it could be argued that the use of peer reports is a form of observation using many observers [Pellegrini, 1996], that such observations are made over a relatively long period of time, and that such reports provide the weighted information of a fairly large number of classmates.

One of the peer-report measures that has attracted considerable attention in recent years is the Participant Role Scales (PRS), developed by Salmivalli et al. [1996, 1998]. They designed a questionnaire to investigate the self and peer estimates of behavior in bullying situations (or participant roles) of 573 Finnish children (12 and 13 years of age). The pupils evaluated on a three-point scale how well each child in their class (including themselves) fit 50 behavioral descriptions. Two items were deleted, but the remaining 48 items were assigned to five different scales: a Bully Scale (descriptions of active, initiative taking, leader-like bullying behavior), an Assistant Scale (more follower than leader-like), a Reinforcer Scale (reflecting tendencies to reinforce the bullying by laughing and inciting), a Defender Scale (supportive of and consoling the victim), and an Outsider Scale (doing nothing, staying out of the bullying situations). Victimization was operationalized by one item: "gets bullied". The authors reported satisfactory internal reliabilities for each of the scales (more than .80 for each scale).

Peer-estimated scores were standardized by class, and each child was identified with a particular (participant) role if their score on that role was both higher than the class mean and higher on that scale than on any of the other scales. A child was identified as a victim if 30% or more of their classmates nominated them as a victim. In this way, they were able to classify 87% of the pupils. The roles of outsider, reinforcer and defender were the most frequent. There were significant gender differences, with more female than male defenders and outsiders, and more male than female bullies, assistants and reinforcers. Salmivalli et al. reported significant but relatively low correlations between self-reports and peer reports. In terms of prevalence, the self-reports were lower on active participation in bullying and higher on defending and being an outsider than the peer reports indicated. Further validation of the scales consisted of investigating the link between participant roles and sociometric status. Bullies were generally rejected or controversial, victims were rejected, and defenders were the most popular. Such links have also been established in other studies [Boulton and Smith, 1994; Lagerspetz et al., 1982; Perry et al., 1988].

In a second study, Salmivalli et al. [1998] investigated the stability of the participant roles over a 2-year period. The researchers now used a shorter version of the PRS, consisting of only 22 items plus the single victim-nomination item. Factor analysis of these items (minus the victim-nomination item) demonstrated the existence of one bullying factor (with the items for bully, assistant and reinforcer all clustering together), and two additional factors for defender and outsider items. Gender differences were again evident. Test–retest correlations indicated a substantial amount of stability for the scale scores (with the exception of defending for the boys and being an outsider for the girls). Stability was lower when a change of classroom environment had taken place, but it was not

clear whether this was due to real changes in a child's behavior or to a change in peer observers.

Sutton and Smith [1999] also employed a shorter version of the original PRS (20 items plus the victimization item), using a nomination procedure for all items (further probing their subjects about whether those nominated behaved "sometimes" or "often" in that way). Because the subjects in Sutton and Smith's study were considerably younger (7–10 years), they interviewed them, while Salmivalli et al. [1996] had simply made use of a self-administered questionnaire. Factor analysis once again indicated the existence of one "bullying" factor plus factors for "defending", "being an outsider" and "victimization". (Unlike Salmivalli et al., who kept the victimization item separate, Sutton and Smith included theirs in the factor analysis.) Scale reliabilities were lower than in the original publication by Salmivalli et al., but they were considered satisfactory. Sutton and Smith set out to adapt the PRS for use with a different sample, but also tried out alternative scoring methods, because they feared that "the standardization of scores by school class could result in children being categorized as having no role due to an extreme score from a classmate" (op. cit. p. 99). Thus, they compared Salmivalli's original criteria of standardization by class (method 1) with three other methods for assigning roles. Method 2 involved no standardization; everyone was given a role corresponding to the highest score received. Method 3 involved standardization by sample and method 4 involved assignment on the basis of the factor analysis. The results based on standardization by class and standardization by sample did not differ. The factor scoring method did not distinguish between bullies, assistants or reinforcers (due to all bullying items forming one factor). Simply assigning on the basis of the highest scale score resulted in all subjects receiving a role. The gender effects were still present in all methods except for the highest scale score method. They refrained, however, from recommending one particular method of assignment and suggested that assigning on the basis of percentage scores might also be a useful alternative. In this paper, we will follow up the latter suggestion.In the Netherlands, Oude Nijhuis [2001] used all 48 original items from the Salmivalli et al. study [1996]. Rather than using a rating system, she employed a nomination procedure (with a maximum of five classmates) and interviewed her subjects ($N = 125$; ages ranged from 8 years, 5 months to 13 years, 8 months) individually. In addition to the PRS, Oude Nijhuis employed the Aggression and Victimization scale, developed by Perry et al. [1988]. Internal reliabilities for the PRS ranged from .74 (assistant; 4 items) to .94 (bully; 10 items), while the alpha for the victimization items (measured separately) was .96 (7 items). She also reported gender differences, in line with those of Salmivalli et al. [1996]. Girls were more often defenders and outsiders, while the boys were more often bullies and assistants. Older children were more often reinforcers. The author also reported a significant relationship between participant role and sociometric status. Bullies were more often rejected, while victims (role assignment based on a *z*-score by class criterion as for the other roles) were more often neglected. Defenders were more often popular.

In short, the PRS has been used by various researchers from different countries (Finland, United Kingdom and the Netherlands), with samples ranging in age from 7- to 14-year-olds, using various procedures (rating plus testing in the classroom, individual testing with nominations plus asking how often, individual testing plus a maximum of nominations), separate formats (the original 50 items, reduced numbers of items, with just one item for victimization or with a set of items for victimization), and with different criteria for assignment. Most studies replicated the gender differences, while both in Finland and in the Netherlands indications for the validity of the PRS were found. Salmivalli et al. [1998] also submitted data demonstrating reasonable stability of roles played in bullying over a 2-year period.

The first aim of our study was to investigate the validity of a shortened and somewhat different version (more items for victimization) of the PRS, which we called the NPRS (N for new) to distinguish it from the original PRS, in a Dutch setting. More specifically, we used a design that is very similar to the one employed by Salmivalli et al. [1996, 1998]. Like them, we collected data on two occasions with a 2-year period in between, and we also collected, on both occasions, data concerning the sociometric status of the children with the various participant roles. Salmivalli and colleagues showed that children both actively and passively involved in the bullying situation tend to be rejected by their peers. We expected to replicate these results.

The second aim of our study was to test a number of different criteria for assigning roles to subjects, since it is our view too that the *z*-score transformation procedure has certain drawbacks. First, by creating *z*-transformations by class, one assumes that every class will have all or most of these roles, but it is also possible for a class to consist only of

bullies and victims, or to have no bullies or victims. We believe the subject of bullying is best approached from an absolute point of view rather than a relative one, as it is unlikely that bullying will be the same in every class. We simply compared classes on the basis of various criteria. A second drawback to assigning roles on the basis of *z*-scores is that someone may be considered a bully with a relatively low albeit positive *z*-score, while another subject may be so considered on the basis of a considerable deviation from the class mean of 0. There is also the Sutton and Smith [1999] argument that the presence of extremely high scorers in a class may mask the presence of other bullies, because their *z*-scores have become low and negative. Third, a relative procedure for assigning roles will make it difficult to use this method to study long-term development or to evaluate the effects of an intervention. We compared the Salmivalli method of assigning roles on the basis of *z*-scores with more absolute percentage criteria which only assign roles if a subject has been nominated sufficiently often by peers. We used three absolute criteria: 10%, 15% and 20% of nominations.

In sum, we had two different aims in mind. First, we sought to assess the psychometric characteristics of the NPRS with a Dutch sample. Second, we wished to compare various ways of assigning roles.

METHOD

The sample

At T1 there were 242 children (123 or 51% boys and 119 girls). The average age of the children was 9 years and 9 months; SD = 8.3; range 100–144 months. At T2 (2 years later) there were 247 children (121 or 49% boys and 126 girls). The average age of the children at that point was 11 years and 8 months; SD = 8.6; range 124–168.

The number of children tested on both occasions was 224. Reasons for the loss of subjects were as follows: children were kept back in a lower grade; some children moved away from the school to a new neighborhood and two children who took part at T1 were withdrawn by their parents, although they were still allowed to act as informants for the NPRS items. We also had 23 new children at T2, because their parents had moved into the catchment area of the schools.

The children were recruited from four different schools. Two of these schools were situated in a medium-sized town (population 100,000) in the vicinity of Amsterdam, one was located in a small town (15,000) in the central part of the Netherlands, while the fourth school was located in a medium-sized town (population 60,000) in the South. Schooling, which is provided free of charge, is available from age 4, and compulsory from age 5. Most children start school at the age of 4. Schools usually serve children from a wide variety of socioeconomic strata, except for some schools in the inner-city areas of large urban conglomerations. The selected schools are thus representative of the wider Dutch population. The percentage of children from ethnic minorities never exceeded 10% in any class. The response rate was high: more than 90% of the parents approached gave permission for their children to participate. The high participation rate is probably due to the active endorsement of the study by all four school directors, and to the fact that the study was predominantly about bullying, a topic parents are concerned about.

Measures

New Participant Role Scales. We reduced the original 48 items as used in the Oude Nijhuis study [2001] to 28 on the basis of their item-total correlations, which are known to resemble factor loadings. Of these 28 items, six covered the initiating leader-like bullies, five their assistants and six the reinforcers. Five items covered the outsiders and six the defenders. This selection of 28 of the original items is very similar to the item set used by Sutton and Smith [1999] and by Salmivalli et al. [1998]. Four additional items were selected to cover the victims. These were selected in the same manner from the seven victimization items of the Aggression and Victimization Scale developed by Perry et al. [1988]. Thus, the NPRS consisted of 32 items in total. Children were asked to nominate peers for each of the 32 items; they could nominate as many peers as they liked, but could not nominate themselves for an item. Children nominated on average 2.5 other children at T1 (range 1.1–4.6) and 1.7 at T2 (range .7–3.0).

Sociometric nominations. We used the procedure proposed by Coie et al. [1982]. Subjects were asked two questions: (1) Which three group members do you like most? (2) Which three group members do you like least? The "most liked" (LM) nominations received by a person and "least liked" (LL) nominations received were standardized by class. Next, the standard scores were transformed into two new variables: SP (social preference; $SP = z_{LM} - z_{LL}$) and SI (social impact; $SI = z_{LM} + z_{LL}$). The resulting scores were standardized as well. Finally, the subjects were assigned to one of five

sociometric status categories as follows: (a) popular, persons with $z_{SP} > 1$, $z_{LM} > 0$ and $z_{LL} < 0$; (b) rejected, persons with $z_{SP} < -1$, $z_{LM} < 0$ and $z_{LL} > 0$; (c) neglected, persons with standardized $z_{SI} < -1$, $z_{LM} < 0$ and $z_{LL} < 0$; (d) controversial, persons with $z_{SI} > 1$, $z_{LM} > 0$ and $z_{LL} > 0$; (e) average, all the individuals not belonging to categories (a), (b), (c) or (d). At T1, there were 23 popular, 31 rejected, 26 neglected, 8 controversial and 154 average children. At T2, the figures were 27, 34, 31, 12 and 143, respectively.

Sociometric ratings. The procedure described above is a relative measure, based on the computation of standardized preference scores and standardized impact scores by class. The sociometric procedure developed by Maassen et al. [1996], on the other hand, is absolute, as it computes probabilities based on the actual judgments given. Both methods result in the assignment of five status groups, but the Coie et al. method does so on the basis of nominations and fixed assumptions about distribution along two dimensions, whereas the method of Maassen et al. does so on the basis of ratings and assumes a chance distribution on the condition of random assignment of scores by the raters, on the basis of which the extremes can be determined (observed minus expected). We used both the traditional method of computing sociometric status, based on nominations and *z*-scores, and the new method, based on rating scales and chance distribution.

Social preference (SP) and social impact (SI), the central variables in the nomination procedures, are transformations of liking and disliking. Maassen et al. [1997] argued that liking/disliking another person at the level of the assessing *individual* should be regarded as a one-dimensional concept. By presenting the assessor with a bipolar rating scale with "like (very) much" and "dislike (very) much" as response anchors of the two poles, and a neutral midpoint, the assessor is able to express his or her affection, repulsion or absence of these feelings towards the assessed person. The concepts "acceptance" and "rejection", traditionally regarded as separate dimensions [e.g. Bukowski et al., 2000], come into play at group level. A computer program is available to split the rating scale into two halves and to calculate the sums of the ratings on the positive and negative halves. Again, these total scores are transformed into two new variables: social preference (which is equal to the score on the complete rating scale) and social impact. These two variables allow classification into the five common sociometric status groups. This method is a probability method and the division into categories is guided by the test of whether social preference, social impact and positive and negative rating totals prove to be higher or lower than expected on the basis of chance. The probabilities are derived from multinomial distributions estimated from the score patterns of the assessors. Thus, this procedure re-introduces an old concept put forward by Moreno [1934] that the opinions of the assessed persons (who coincide with the group of assessors) about their group should also be taken into account. We used a conventional (5%) level of statistical significance. At T1, we had 61 popular, 58 rejected, 5 neglected, 0 controversial and 118 average children ($N = 241$; 1 missing). For T2, these figures were 71, 54, 17, 3 and 102, respectively.

Procedure

Children were tested individually in a quiet room in their school by one of nine female assistants, who were unfamiliar to them. We used two sessions for these tests. In session 1, we administered the NPRS. Children were provided with a list of names of classmates (at T1), and were asked to nominate peers (unlimited nominations) who fitted the description of the items. If they did not mention anybody, they were prompted once. In the second session, we administered the two sociometric procedures described above. For the rating procedure, we used a 7-point scale at both points in time.

Children were given instructions about the various procedures and were told that all information would remain confidential and would not be passed on to any of their peers. This procedure was repeated 2 years later.

RESULTS

Principal Component Analyses at T1 and T2

We submitted the NPRS raw nominations (T1), divided by $N-1$ (assessors) to a principal component analysis with varimax rotation. The Scree plot pointed to the existence of four factors, one large factor for bullying and three smaller ones for outsider, defender and victim. Eigenvalues associated with these factors were 11.2, 4.9, 3.5 and 3.5, respectively, for a total of 72.3% explained variance. A maximum likelihood analysis with oblimin rotation resulted in approximately the same results. We ran the same analyses with the T2 data, with very similar results, that is, the same four factors, and the same composition of items; total variance explained

now amounted to 78.3%. Eigenvalues were now 10.8, 6.1, 4.5, and 3.6, respectively. We deleted four items from the set, that is, item 4 (''Someone who threatens to go tell the teacher if the others do not stop bullying''), item 11 (''Someone who tells the others it does not make sense to join in the bullying''), which had low loadings throughout and items 25 (''Someone who is always there, even if s/he does not do anything'') and 26 (''Someone who comes over to see what is happening''), which had considerably lower loadings at T2 than at T1 in both types of factor analysis. The items also had multiple loadings at T2. (Tables with the various results are available from the authors on request.)

Confirmatory Factor Analyses

Because the factor solution came up with one large bullying factor (not allowing for a distinction between active initiative-taking bullying and less active joining in later), and because we considered the distinction between the bully on the one hand and the follower on the other conceptually relevant, we decided to compare two models, using confirmatory factor analyses. The first model was the 4-factor model, provided by the principal component analysis and the second one was a 5-factor model, with the large bullying factor divided up into a bully and a follower role, allowing for a correlation between the two latent factors. We wanted to know whether the 5-factor model would fit the data better than the 4-factor model.

Several fit indices are available to determine the goodness-of-fit of the two models. The first index is the chi-square test. This test compares the covariance matrix of the model with the covariance matrix of the sample. The chi-square test is sensitive to the number of variables in the model and to the sample size. Several fit indices were developed to avoid these problems. The comparative fit index (CFI) [Bentler, 1990] estimates the fit of a model with respect to a continuum that goes from the null model (no relationship is estimated among the variables) to the saturated model (all the possible relationships between the variables are estimated). The CFI is bounded by 0 and 1, and a model is considered acceptable when it shows a value higher than .90.

The relevant statistics were as follows. At T1 for the 4-factor model, the figures were as follows: χ^2 (df = 344) = 937.45; $P < .0001$; CFI = .92. For the 5-factor model they were as follows: χ^2 (df = 340) = 861,9; $P < .0001$; CFI = .93. The $\Delta\chi^2$ (df = 4) = 75.55; $P < .0001$. The zero-order correlation coefficient between the two latent variables ''bully'' and ''follower'' was .96. At T2 the relevant statistics were as follows. For the 4-factor model: χ^2 (df = 344) = 1297.71; $P < .0001$; CFI = .89. For the 5-factor model these figures were: χ^2 (df = 340) = 1228.71; $P < .0001$; CFI = .90. The $\Delta\chi^2$ (df = 4) = 68; $P < .0001$. The zero-order correlation coefficient between the latent variables ''bully'' and ''follower'' was .96. Thus, in both cases, the 5-factor solution provided a better fit than the 4-factor solution, although it can be argued that the 4-factor solution also met the relevant criteria.

Reliability of the Scales

Internal consistency coefficients (Cronbach's alpha) were as follows: bully (6 items) .96 and .95 at T1 and T2, respectively, follower (8 items) .95 and .94, outsider (6 items) .92 and .96, defender (4 items) .84 and .92 and victim (4 items) .91 and .94.

At T1, the correlations between the scales were as follows: between bully and follower .93, between bully and outsider −.43, between bully and defender −.26 and between bully and victim .16 (all P's < .01). Follower and outsider correlated −.43, follower and defender −.25 and follower and victim .15 (all P's < .02). Between outsider and defender, the correlation was .6 ($P < .001$), while outsider and victim correlated −.03 (n.s.). Defender correlated −.13 with victim ($P < .03$; all $N = 240$).

At T2, the correlations between the scales were as follows: between bully and follower .90 ($P < .001$), between bully and outsider −.17 ($P < .01$), between bully and defender −.03 (n.s.) and between bully and victim .04 (n.s.). Follower and outsider correlated −.18 ($P < .01$), follower and defender −.07 (n.s.) and follower and victim .03 (n.s.). Between outsider and defender, the correlation was .67 ($P < .001$), while outsider and victim correlated −.20 ($P = .001$). Defender correlated .15 with victim ($P < .02$; all $N = 241$). In general, correlations between the various scales decreased from T1 to T2 (with the exceptions of that between defender and outsider, and between defender and victim).

Distributions of the Various Roles According to Varying Criteria

In Table I we present the numbers of bullies, followers, outsiders, defenders, victims, children uninvolved and children not classifiable according to four criteria: (1) z-scoring and application of the original Salmivalli et al. classification rules, also for victims. This means that roles can only be assigned if the z-scores are above zero, if classification is based on the highest z-value, if z-values which differ less

TABLE I. Numbers of Bullies, Followers, Outsiders, Defenders, Victims, Children Not Involved and Unclassifiable by four Criteria: *z*-Scoring, 10%, 15% and 20% Scoring, at T1 and T2

	Bullies *N* (%)	Followers *N* (%)	Outsiders *N* (%)	Defenders *N* (%)	Victims *N* (%)	Not involved *N* (%)	Unclassified *N* (%)
T1 criterion							
z	22 (9)	40 (16)	52 (21)	48 (19)	35 (14)	32 (13)	13 (5)
10%	35 (14)	14 (5)	78 (32)	37 (15)	24 (9)	44 (18)	10 (4)
15%	29 (11)	10 (4)	57 (23)	21 (10)	21 (9)	95 (39)	9 (1)
20%	23 (9)	1 (0)	42 (17)	14 (5)	18 (7)	142 (58)	2 (.5)
T2 criterion							
z	25 (10)	42 (17)	49 (19)	42 (17)	31 (12)	43 (17)	5 (6)
10%	16 (6)	14 (5)	41 (16)	29 (11)	26 (10)	115 (46)	6 (2)
15%	14 (5)	9 (3)	26 (10)	23 (9)	21 (8)	151 (61)	3 (1)
20%	10 (4)	7 (2)	20 (8)	19 (7)	15 (6)	174 (70)	2 (.5)

than .1 from each other lead to an assignment as not classifiable, and if those who have no *z*-values above zero are assumed not to be involved in bullying. For the bully versus follower distinction, we simply assigned subjects on the basis of the highest score, even if the difference between the two *z*-scores was less than .1; (2) classification of roles if nominated by at least 10% of the peers. Here we used an average criterion; that is to say, assignment could only take place if on average (over the various items comprising a scale) the subject had been nominated by at least 10% of their classmates. For instance, in a class of 21 pupils, one needed 6 (items of the bully scale) times 2 (10% of 20) = 12 nominations in order to meet the 10% criterion for bully. Differences of 1% or less led to an assignment as not classifiable; (3) classification if nominated by at least 15% (on average; 18 nominations in a class of 21 pupils) and (4) classification if nominated by at least 20% (on average; 24 nominations in a class of 21 pupils). The data for T1 are presented at the top and those for T2 at the bottom.

Inspection of Table I reveals considerable differences between the various classification criteria. First, if we use the *z*-criterion we find that bullies are accompanied by a rather large number of followers (22 bullies and 40 followers at T1 and 25 bullies and 42 followers at T2). Using a percentage criterion, it would appear to be the other way around. There are now more bullies than followers. At T1, the number of bullies is higher when using a percentage criterion compared to the *z*-criterion classification, but this is not the case at T2. Second, whatever criterion used, outsiders and those not involved would appear to make up the majority of the subjects. There are more outsiders than those not involved using the *z*-criterion (and also using the 10% criterion at T1), but the not-involved classification becomes the largest using all other percentage criteria, both at T1 and at T2. Third, generally speaking, the *z*-criterion assigns more subjects to one of the five roles defined (bully, follower, outsider, defender and victim), indicating some type of involvement in bullying, and fewer subjects to the not-involved classification, while the percentage criteria all assign fewer subjects to any of these five roles and more to the not-involved classification. Finally, applying the *z*-criterion makes more subjects unclassifiable than applying any of the percentage criteria. It should be borne in mind that the criteria for categorizing someone as not classifiable are not fully comparable for the different methods. Furthermore, the *z*-criterion tends to assign about equal percentages to each role at each point in time, while the numbers appear to differ more with each and every one of the percentage criteria. Finally, there is also a difference in trends as indicated by the criteria. The *z*-criterion does not seem to signal a trend at all, as the difference between T1 and T2 is never more than 2%. The percentage criteria would appear to be clearer in this respect; they all indicate a substantial decrease in the number of bullies, outsiders and defenders (except in the case of the 20% criterion), while the numbers of followers (except in the case of the 20% criterion) and victims would appear to be stable.

Agreement Between the Various Criteria

We computed kappa coefficients to assess the agreement between the various criteria used for classification. At T1 the kappa coefficients were as follows: .60 for agreement between the *z*-criterion and the 10% criterion, .47 between the *z*-criterion and the 15% criterion and .30 between the *z*-criterion and the 20% criterion. At T2, these figures

were .48, .39 and .30, respectively. Kappa coefficients in the .30–.50 range are considered to be moderate; those higher are considered indicative of a strong correlation. All P's were $<.001$.

Test–Retest Stability at the Level of Classifications

We also computed kappa coefficients to assess agreement between the T1 and the T2 classifications for each of the criteria used. The coefficients were as follows: for the z-criterion $\kappa = .33$, for the 10% criterion $\kappa = .24$, for the 15% criterion $\kappa = .33$, for the 20% criterion $\kappa = .34$. These figures indicate moderate to low stability of the classifications over a 2-year period. We also checked whether children had changed classes, and had thus been confronted with new peers, but this was only the case for eight children. The composition of the classes was therefore very stable. Coefficients for the raw scale scores, divided by $N-1$, were also computed. These were .74 (bully), .61 (follower), .65 (outsider), .53 (defender) and .55 (victim).

Other Validity Indices

Gender differences. Both Salmivalli et al. [1996] and Sutton and Smith [1999] reported gender differences, with boys being more often bullies and followers, while girls were more often defenders and outsiders. We checked whether this was the case for our criteria. We ran chi-square analyses with sex and the classifications (bully, follower, outsider, defender, victim, and not involved) as factors for each of the four criteria. At T1, we found that using the z-criterion and the 10% and 15% criteria the results were all the same. Girls were more often outsiders and defenders, while boys were more often bullies and followers. Applying the 20% criterion, boys were no longer found to be followers more often than girls (there was only one; see Table I), but the other links with gender were still present. Chi-squares ran from 33.1 to 47.9 and LR chi-squares ran from 38.1 (20% criterion) to 55.5 (z-criterion; all P's $<.001$). Adjusted standardized residuals (best seen as the standardized measure for observed minus expected per cell) ranged from 2.3 (more female defenders according to the 10% criterion) to 5.8 more boy followers (z-criterion). At T2 the results were similar. Girls were still more often defenders, while boys were still more often bullies and followers. This was found for all criteria used. For the z-criterion, we also found an association between gender and victimization. Girls were more often victims. Chi-squares ran from 23.3 (20% criterion) to 32.5 (z-criterion), and LR chi-squares ran from 28.6 (10% criterion) to 34 (z-criterion). Adjusted standardized residual z-values ranged from 2.1 (more female victims according to the z-criterion) to 4.2 more boy followers (z-criterion; all P's $<.001$).

Links with Sociometric Status

We ran chi-square analyses with sociometric status (popular, rejected, neglected, controversial and average according to the nomination-based method, and according to the ratings-based method) and the classifications as factors for each of the four criteria, and for both T1 and T2. Since in most of these analyses the number of cells with expected values less than 5 was rather large, we need to be cautious with our interpretation. An overview of the results can be found in Table II, where we present the results of 16 chi-square analyses (4 criteria $\times$ 2 methods of computing sociometric status by two points of measurement). To make this table more accessible, we separated the positive standardized residuals (those above or equal to 2) from the negative ones (those below or equal to -2). Thus, in the first cell of Table II (with the marginals "Z" and "popular"), we see that using the z-way of computing bullying roles, we found that defenders were more often popular, both when using the nominations-based method and when using the ratings-based method, and that this not only obtained at T1, but also at T2. To complete the description of that cell, we can also see that bullies were more often popular than expected by chance at T2 (nominations-based method) and that the outsiders were also more often popular than expected at T1 (ratings-based method).

We start with the results of the nominations-based sociometric statuses. At T1, we found a link between bullying roles (whatever the criterion used) and sociometric status. Chi-squares ran from 57.9 (z-criterion) to 88.2 (15% criterion), and LR chi-squares ran from 58.1 (z-criterion) to 70.6 (15% criterion; all P's $<.001$). In general, prosocial behavior as demonstrated by defenders was associated with popular status, while aggressive behavior as demonstrated by bullies, who have been known to be both proactively and reactively aggressive, and by victims, who have been known to be reactively aggressive, was associated with rejected status. Nevertheless, the majority of the roles were found among the average status children, except for defenders and victims. In all cases, the percentage of defenders or victims who were average was less than 50%.

TABLE II. Distribution of Residual Values (Positive and Negative) Across Five Sociometric Status Groups (Nominations and Ratings-Based Method) at T1 and T2

	Popular		Rejected		Neglected		Controversial		Average	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Positive residual values (2 or more)										
z	Def	Def	Bully		Not	Out		Bully		Follow
		Bully	Vict	Vict						
	Def	**Def**	**Bully**						**Follow**	
	Out		**Vict**	**Vict**	**Vict**	**Out**		**Bully**	**Not**	**Not**
10%	Def	Def	Vict	Vict		Not		Bully		
	Def		**Bully**	**Bully**		**Out**		**Bully**	**Follow**	
	Out		**Vict**	**Vict**				**Follow**	**Not**	
15%	Def	Def	Bully		Not	Not		Bully		Not
			Vict	Vict						
	Def	**Def**	**Bully**	**Bully**						
	Out		**Vict**	**Vict**		**Out**		**Follow**	**Not**	
20%	Def	Def	Bully		Not	Not		Bully		Not
	Follow		Vict	Vict						
	Def		**Bully**	**Bully**						
	Out		**Vict**	**Vict**		**Out**		**Follow**	**Not**	
Negative residual values (−2 or less)										
z	Bully		Out	Out						Bully
				Def						Vict
	Follow	**Follow**	**Out**	**Out**					**Def**	
	Vict	**Vict**	**Def**	**Def**					**Vict**	**Vict**
				Not						
10%			Def			Vict				Bully
			Not	Not					Vict	Vict
	Bully	**Bully**	**Out**	**Out**					**Vict**	**Vict**
	Vict	**Vict**	**Def**	**Def**						
	Not									
15%			Not	Not		Vict	Not		Vict	Vict
	Bully		**Out**	**Out**						
	Vict	**Vict**	**Def**	**Def**					**Def**	
	Not		**Not**	**Not**					**Vict**	**Vict**
20%			Not	Not		Not				Bully
									Vict	Vict
	Bully		**Out**	**Out**					**Def**	
	Vict	**Vict**	**Def**	**Def**					**Vict**	
	Not		**Not**							

Follow, follower; Out, outsider; Def, defender; Vict, victim.
Names and abbreviations in bold indicate results from the ratings-based method.

At T2, we found roughly the same associations; that is, defenders were still popular, victims were still rejected, but the bullies were now controversial (all criteria). Those not involved were still predominantly neglected. Fewer children now had an average status. Chi-squares ran from 81.8 (20% criterion) to 105.5 (*z*-criterion), and LR chi-squares ran from 65.6 (20% criterion) to 95.1 (*z*-criterion; all P's $< .001$).

At T1 with the ratings-based sociometric status data, we found more or less the same associations. Defenders were popular, irrespective of the criterion used, and so were the outsiders, while bullies were most often rejected, and victims were always rejected. The only role which consistently had predominantly an average status was that of those who were not involved. Chi-squares ran from 117.7 (20% criterion) to 141.6 (15% criterion), and LR chi-squares ran from 114.1 (20% criterion) to 145 (*z*-criterion; all P's $< .001$).

At T2, we found the following. Chi-squares ran from 80 (20% criterion) to 121.4 (*z*-criterion), and LR chi-squares ran from 71.6 (20% criterion) to 122.3 (*z*-criterion; all P's $< .001$). Victims were more often rejected (all criteria), while bullies were more often rejected according to the higher percentage

criteria, but more often controversial according to the *z*-criterion, and the 10% criterion. Please bear in mind that there were only three controversial children (two of whom were bullies) at T2, as computed with the ratings-based method. Defenders were more often popular (*z*-criterion and 15% criterion). Outsiders were more often neglected (all criteria, while earlier on they had been popular). Looking at the links between role in bullying and sociometric status for each gender separately, we found that victims remained rejected throughout, and defenders were still popular, at least among the girls. The sociometric status of bullies became more diffuse.

There were also differences between the results found with the two methods of computing sociometric status. While defenders were popular both at T1 and T2 with the nominations-based method, these links were less strongly present with the ratings-based method. That is, defenders were always more often popular than could be expected by chance at T1, but at T2 this link was only statistically significant with the *z*-criterion and the 15% criterion. In addition, with the nominations-based method bullies were more often found to be controversial at T2, and rejected at T1, while with the ratings-based method they were invariably more often rejected at both T1 and T2, but only controversial when we used the *z*-criterion and the 10% criterion. Those not involved were mainly neglected with the nominations-based method, but no such associations were found with the ratings-based method. Outsiders were popular at T1 (ratings-based method), but neglected at T2, while such associations were absent with the nominations-based method.

Comparing the Various Ways of Assigning Roles

Secondary roles. Since each of the criteria aims at assigning the children to a certain role, and does so on the basis of high scores (either *z*-scores or percentages), it is possible for an individual to be given a certain classification despite relatively high scores for another classification. In other words, an individual could be classified as a bully because of (a) meeting the minimum requirement for that category and (b) having the highest score for that category. Yet, it was still possible for the same person to meet the criterion for another category (a follower or a victim), but not to be classified as such because of a lower score on that scale than on the one for bully. We compared the various approaches on secondary roles. The results are presented in Table III.

Inspection of Table III shows that secondary roles are more characteristic of *z*-scores and the 10% criterion than of the two other criteria. It can also be seen that these secondary roles are usually similar to the role finally assigned. Most often children with a high bully score, but not classified as such, were classified as followers, and vice versa. Most children with a high outsider score who did not make the criterion for outsider were usually classified as defenders, and vice versa. Finally, there would appear to be children with a high victim score who were classified as bullies or followers. There were also children classified as victims whose secondary role was that of a bully or follower. Their number decreased from the 15% criterion up, and this also appeared to be less the case at T2 than at T1. This points to the existence of bully/victims, a classification that has so far been neglected within the original classification system. Yet, it would be possible to create such a category by changing the rules for assignment.

Comparisons at class level. We also inspected the distributions of the various roles (bully, follower, outsider, defender and victim) per class, employing each of the various criteria, and at both T1 and T2. The results are shown in Table IV.

It can readily be seen that using the *z*-criterion, one will always find representatives of each role in every class. This is much less the case when one uses absolute (percentage) criteria. In fact, the higher the percentage criterion, the greater the number of classes without representatives for at least one role. We also checked whether there were classes with no representatives for more than one role. When using the 10% criterion, this was the case for only one class at T1, but not at T2. In the case of the 15% criterion, we found two classes with two roles missing at T1, and three classes at T2. We also found one class with three roles missing. Using the 20% criterion we found four classes with at least two roles missing at T1, and three classes with two roles missing at T2, in addition to one class with three roles missing and another class with as many as four roles missing. We would like to remind the reader that the sample consisted of only nine classes. These figures, therefore, are tentative at best.

DISCUSSION

The first aim of this study was to check the reliability and, in particular, the validity of a new

TABLE III. Overview of "Secondary Role" Scores for Every Classification, and Their Corresponding Classifications at T1 and T2

	High on bully score, but not classified as such	High on follower score, but not classified as such	High on outsider score, but not classified as such	High on defender score, but not classified as such	High on victim score, but not classified as such
T1 criterion					
z	18 Follower	17 Bully	18 Defender	1 Bully	6 Bully
	4 Defender	1 Outsider	9 Victim	13 Outsider	10 Follower
	2 Victim	2 Defender	7 Not classifiable	2 Victim	2 Outsider
		3 Victim		3 Not classifiable	2 Defender
		1 Not classifiable			3 Not classifiable
10%	5 Follower	19 Bully	17 Defender	4 Bully	6 Bully
	1 Defender	1 Outsider	5 Victim	23 Outsider	1 Follower
	3 Victim	1 Victim	4 Not classifiable	2 Victim	2 Outsider
	2 Not classifiable			4 Not classifiable	1 Not classifiable
15%	2 Victim	18 Bully	10 Defender	1 Bully	3 Bully
	1 Not classifiable	1 Outsider	1 Not classifiable	7 Outsider	1 Not classifiable
		1 Victim		1 Not classifiable	
20%	1 Victim		Defender (5)	3 Outsider	3 Bully
	1 Not classifiable		5 Defender	1 Not classifiable	1 Not classifiable
T2 criterion					
z	13 Follower	15 Bully	2 Follower	4 Bully	2 Bully
	2 Defender	2 Defender	14 Defender	2 Follower	3 Follower
	4 Victim	5 Victim	4 Victim	6 Outsider	2 Outsider
	2 Not classifiable	2 Not classifiable	5 Not classifiable	2 Victim	2 Defender
				10 Not classifiable	4 Not classifiable
10%	2 Defender	9 Bully	1 Follower	2 Bully	2 Outsider
	2 Victim	1 Outsider	15 Defender	12 Outsider	1 Not classifiable
		1 Defender	4 Victim	2 Not classifiable	
		2 Victim	5 Not classifiable		
15%		3 Bully	6 Defender	1 Bully	2 Outsider
			3 Victim	5 Outsider	
			2 Not classifiable	1 Not classifiable	
20%		3 Bully	4 Defender	5 Outsider	2 Outsider
			2 Victim	1 Not classifiable	
			1 Not classifiable		

TABLE IV. Distributions (Ranges) of the Various Roles (Bully, Follower, Outsider, Defender and Victim) Identified by Class for Each of the four Criteria, Plus the Number of Classes Without Representatives for That Role

	Range and (*N* of classes without)				
Role	Bully	Follower	Outsider	Defender	Victim
T1 criterion					
z Scores	1–4	2–7	1–9	2–10	2–6
10%	0–6 (1)	0–3 (2)	5–14	0–10 (1)	1–4
15%	0–5 (1)	0–3 (2)	2–12	0–8 (2)	1–4
20%	0–4 (1)	0–1 (8)	1–10	0–7 (3)	0–4 (1)
T2 criterion					
z Scores	1–5	2–9	3–10	3–7	1–5
10%	1–3	0–8 (4)	1–9	0–11 (1)	2–5
15%	0–3 (2)	0–6 (6)	0–8 (1)	0–8 (2)	0–5 (1)
20%	0–3 (3)	0–5 (7)	0–8 (2)	0–6 (2)	0–4 (1)

version of the PRS. The data indicate that this new version can be considered reliable and valid. That is, its scales were internally consistent, and the test–retest reliability over a 2-year period was reasonably high at the level of the scale scores. There was a moderate degree of test–retest stability at the level of the classifications. In addition, the classifications were linked to gender and to sociometric status in a manner that replicated the original results of Salmivalli et al. [1996]. Moreover, the CFA demonstrated that a 5-factor solution fitted the data best. We found with a CFA that a 4-factor solution had a lower, albeit acceptable, fit than the 5-factor solution which distinguished between bullies and followers. We also found that there was a considerable degree of overlap between being a bully and being a follower.

On the other hand, we now have data to support the distinction between bullies and followers. Both Salmivalli et al. [1998] in their shorter version of the PRS with somewhat older children (14–15 years old), and Sutton and Smith [1999] in their version of the PRS with younger children (7–10 years old) only carried out exploratory factor analyses, pointing to

one large pro-bullying factor. The solution chosen by us and by Sutton and Smith was to combine the items for assistant and reinforcer into one large "follower" classification, because we wished to distinguish for conceptual reasons between active, initiative-taking bullies and those who join in at a later stage. The data produced by this study demonstrated convincingly that the distinction between bullies and followers is valid. We not only have the results of the CFA to back up the distinction, but also the links with sociometric status, which were different for bullies and followers. Sutton et al. [1999] demonstrated that bullies had a better theory of mind than followers.

We found, using the nomination-based method of computing sociometric status, that bullies were more often rejected (T1), more often controversial (T2) and less often average (T2), while no such associations were found for followers. For the ratings-based way of assigning sociometric status, we again found the bullies to be more often rejected (T1 and T2) and less often popular (T1, except for the *z*-scoring at T1 and T2), while such associations were not or were scarcely found for followers. Followers and bullies were both more controversial at T2 (10% criterion), and followers were more often average (T1, *z* and 10% criteria). Thus, followers did demonstrate different links to sociometric status than bullies, but we need more evidence to support the distinction between the two. One obvious way to further validate the distinction would be by way of systematic observation.

At T2 bullies were more often controversial, using either the nominations-based method or the ratings-based method. With the latter method, followers were also more often controversial. One explanation could be that bullies (and to a lesser extent followers, but bear in mind their relatively low numbers at T2) who still bully at the end of their primary school career are excellent leaders, who use a double strategy to get their way. Hawley [1999] claims that being bi-strategic, that is, both coercive and prosocial in one's dealings with peers, is a very successful way of getting to the top of the dominance hierarchy. Since prosocial behavior is evaluated by peers in a positive manner, while coercive behavior is evaluated negatively, it is possible that this is reflected in a controversial status, characterized by many likes and many dislikes. This finding needs to be studied in more detail before we can say more of the sociometric status of children with bi-strategic ways of dealing with peers. In addition, we should not lose sight of the fact that with the ratings-based method, bullies were also more often rejected. That the link between bullying and controversial status at T2 was stronger for one method (nominations) than the other (ratings) may be a result of the fact that the ratings-based method assigns fewer children to a controversial status (i.e. 3 versus 12).

As to the identification of bullies and followers, this could be done in a different manner. One way to better identify followers and distinguish them from bullies might be to ask more explicitly whether someone who has been nominated for a certain bullying item is the one who starts the behavior or is the one who joins in once someone else has started.

Other links between roles played in bullying and sociometric status do support the validity of the NPRS. Defenders, who behave in a prosocial and empathic way, were predominantly high on popularity (at T1 and T2 according to the nomination-based method and at T1 according to the ratings). They were also less often rejected, especially according to the ratings-based method, at both T1 and T2. Victims were rejected, irrespective of time of measurement or method employed to compute sociometric status. The links between being an outsider and sociometric status and between not being involved in bullying and sociometric status were less evident according to the traditional nomination-based method than to the ratings-based method. With the latter method, outsiders fell more often in the neglected category (T2) and were less often rejected (T1 and T2). Those not involved were predominantly average (T1). With the nomination-based method, it was especially those not involved who were more often neglected by the peers.

Our NPRS distribution of roles was characterized by gender differences. We found boys to be more often bullies and followers, while girls were more often outsiders and defenders. While these findings have also been reported by Salmivalli et al. [1996], and thus can be seen as replications, this may also point to a certain lack of content validity in the NPRS, as some of the more indirect ways of bullying as practiced by girls may be insufficiently covered by the items it uses. We are thinking here of such bullying behaviors as "damaging someone's reputation" or "social exclusion". What is more, we also need suitable descriptions for followers in this vein. Vice versa, it may also be that more typically male ways of defending ("jumping in between on behalf of the victim", "physically or verbally attacking the bully on behalf of the victim") need to be included. Being an outsider, someone who deliberately seeks to avoid bullying situations, may be more typical of girls, who are simply more prone

to shun violence as displayed by boys. Alternatively, it may be that even with the inclusion of more female bullying items and more male defending items, we will still find gender differences, simply because boys do more bullying and girls behave in more prosocial ways and will be more likely to defend victims (by saying it is no use bullying or by threatening to tell the teacher). In any case, being a defender would appear to be more highly appreciated by girls than by boys, since defenders were popular amongst girls and not amongst boys. These gender differences (including the gender differences in the links between certain bullying roles and sociometric status) bring to light the issue of what boys and girls see as bullying and how they evaluate this. There may not only be cultural differences [Smith et al., 2002] in what is considered to be bullying, but also gender differences as well as differences in what is acceptable or not. Researchers usually employ a definition of bullying which ensures that the participants know what is meant by bullying, but it might also be advisable to use a list of the various ways of bullying (verbal, instrumental, physical, relational) to ensure that every subject interviewed has the same understanding of the concept of bullying.

The second aim of our study was to establish whether a relative measure (such as *z*-scores) or an absolute measure (such as percentage scores) is the best way to assign roles. While the results indicated that percentage scores can be used validly, it is not possible to conclude that we should only use percentage scores. Much depends on the goal of the research. For instance, if we wish to identify developmental trends, it may be more advisable to use percentage scores, as they seem to indicate—in line with other research with self-reports [Smith et al., 1999]—that bullying decreases with age. Salmivalli et al. [1996, 1998] used both a relative criterion for all roles except the victim role, and an absolute criterion to identify the victim (nominated as such by at least 30%). Using that absolute criterion she reported a decrease in victimization. We also found a decrease in victimization with our somewhat younger group, both for the *z*-criterion and for the percentage criteria of 15% and 20%. Similarly, it may be more useful to use percentage scores if we wish to evaluate the results of an intervention program, as we need to apply the same criteria both before and after the intervention. Applying an assignment strategy based on *z*-scores would simply lead to a similar distribution as before the intervention, which might erroneously lead to the conclusion that the intervention had not effected a change, when in fact it might have done. Since the *z*-score assignment procedure appears to assign more children to a role indicative of some involvement in bullying, this procedure might come in handy when we wish to select subjects for inclusion in a treatment program. We will discuss some of the more pertinent findings concerning the comparison between relative and absolute criteria below.

First, the concordance coefficients indicated that classifications on the basis of *z*-scores resemble most the classifications on the basis of 10%, and become less similar with increasing percentage criteria. This was the case at both T1 and T2. This means that with absolute criteria quite a few children are assigned differently than when the relative criterion is applied. One difference was that the *z*-scoring method identified more followers and fewer bullies, while the percentage scoring identified more bullies and fewer followers. To what extent one method is superior to the other cannot be determined on the basis of these results alone. The relative criterion seems to assign more children to the victim role, but it must be borne in mind that Salmivalli et al. [1996, 1998] did not use that criterion for victimization. In fact, with the relative criterion, one will find representatives of every role in every class, while when using an absolute criterion, one will find classes without bullies or any of the other roles identified. Thus, using the percentage criteria one usually obtains larger differences between the classes themselves than with the *z*-criterion. Without a "gold standard" for identifying children as bullies or followers, etc., we have no absolutely certain way to assess which method is the best, but it would seem unlikely that all of these roles can always be identified in every class. The absolute method identifies fewer subjects who cannot be classified, and more subjects as not involved. While the first may be seen as an advantage, this cannot be claimed for the second. It may be that the *z*-scoring produces too many false positives, while the percentage scores produce too many false negatives.

As to the validity data, each criterion would appear to show the same gender differences, both at T1 and at T2. And more or less the same links were found between classification and sociometric status, irrespective of the criterion used, and irrespective of the method for computing sociometric status used. While none of these results seem to support the use of one method or the other, it should also be obvious that each criterion can lay claim to some validity. A case for the use of a higher absolute criterion comes from the data pertaining to secondary roles, which seem to indicate that both the

z-criterion and the 10% criterion identify large numbers of secondary roles, while these decrease considerably for the 15% and 20% criteria, both at T1 and at T2. We are tempted to advise the use of the 15% criterion on the basis of a number of arguments and outcomes. One, percentage scores are absolute criteria and may be better able to uncover differences between groups or times. Two, the use of a percentage criterion may make it easier to classify subjects. Three, the data presented here provide indications for the validity of percentage criteria too. Four, classification on the basis of the higher percentage criteria goes with fewer secondary roles. This makes the 15% and 20% criteria viable alternatives to *z*-scoring. Further validation of the criteria may be undertaken by collecting other data on role behavior, such as teacher reports, self-reports and, in particular, systematic observation. With the use of advanced statistical techniques such as Receiver Operating Characteristics [Swets, 1996] we may be able to develop a viable "gold standard".

Salmivalli et al. [1996, 1998] used ratings instead of nominations. We are much in favor of ratings, because one can collect the judgments of all concerned on all concerned, while nominations (even unlimited nominations) by necessity cover fewer people. Our reason for not using ratings for the NPRS was of a practical nature. It would require tremendous effort on the part of our subjects, and would make the procedure for collecting the data unnecessarily long and tedious. Nominations are easier to carry out, but have the disadvantage that they may be sensitive to an experimenter-effect. In other words, an experimenter who is more insistent might elicit more names (nominations) than one who is easily satisfied. Absolute percentage scores would then differ according to the experimenter, while such effects would not occur with *z*-scores. A solution would be to randomly assign experimenters to classes. Alternatively, one could reduce the number of items and try ratings.

Finally, we were predominantly concerned with the assignment of roles, based on the answers provided to the various items. We should not, however, forget that the NPRS also allows the use of continuous scales for each of the roles, and that these scales were consistently reliable and highly stable over a 2-year period.

REFERENCES

Ahmad Y, Smith PK. 1994. Bullying in schools and the issue of sex differences. In Archer J (ed): "Male Violence." London: Routledge, pp 70–83.

Alsaker FD, Valkanover S. 2001. Early diagnosis and prevention of victimization in kindergarten. In Juvonen J, Graham S (eds): "Peer Harassment in School. The Plight of the Vulnerable and Victimized." New York: Guildford, pp 125–144.

Bentler PM. 1990. Comparative fit indexes in structural models. Psychol Bull 107:238–246.

Björkqvist K, Lagerspetz K, Salmivalli C, Kaukiainen A. 1992. Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. Aggr Behav 18:117–127.

Boulton MJ, Smith PK. 1994. Bully/victim problems in middle school children: Stability, self-perceived competence and peer acceptance. Br J Dev Psychol 12:315–329.

Bukowski WM, Sippola L, Hoza B, Newcomb AF. 2000. Pages from a sociometric notebook: An analysis of nomination and rating scale measures of acceptance, rejection and social preference. In Cillissen AHN, Bukowski WM (eds): "Recent Advances in the Measurement of Acceptance and Rejection in the Peer System." San Francisco: Jossey-Bass, pp 11–26.

Camodeca M, Goossens FA, Meerum Terwogt M, Schuengel C. 2002. Bullying and victimization among school-age children: Stability and links to proactive and reactive aggression. Soc Dev 11:332–345.

Coie JD, Dodge KA, Coppotelli H. 1982. Dimensions and types of social status: A cross-age perspective. Dev Psychol 18:557–570.

Crick NR, Bigbee MA. 1998. Relational and overt forms of peer victimization: A multi-informant approach. J Clin Consult Psychol 66:337–347.

Hawley P. 1999. The ontogeny of social dominance: A strategy-based evolutionary perspective. Dev Rev 19:97–132.

Juvonen J, Graham S. 2001. "Peer Harassment in School. The Plight of the Vulnerable and Victimized." New York: Guildford.

Juvonen J, Nishina A, Graham S. 2000. Peer harassment, psychological adjustment, and school functioning in early adolescence. J Educ Psychol 92:349–359.

Kochenderfer BJ, Ladd GW. 1997. Victimized children's responses to peers' aggression: Behaviors associated with reduced versus continued victimization. Dev Psychopathol 9:59–73

Lagerspetz KMJ, Björkquist K, Berts M, King E. 1982. Group aggression among school children in three schools. Scand J Psychol 23:45–52.

Maassen GH, Akkermans W, van der Linden JL. 1996. Two-dimensional sociometric status determination with rating scales. Small Group Res 27:56–78.

Maassen GH, van der Linden JL, Akkermans W. 1997. Nominations, ratings and the dimensions of sociometric status. Int J Behav Dev 21:179–199.

Menesini E, Eslea M, Smith PK, Genta ML, Gianetti E, Fonzi A, Costabile A. 1997. Cross-national comparison of children's attitudes towards bully/victim problems in school. Aggr Behav 23:245–257.

Moreno JL. 1934. "Who Shall Survive? A New Approach to the Problem of Human Interrelations." Washington, DC: Nervous and Mental Disease Publishing Co.

Morita Y, Soeda H, Soeda K, Taki M. 1999. Japan. In Smith PK, Morita Y, Junger-Tas J, Olweus D, Catalano R, Slee P (eds): "The Nature of School Bullying: A Cross-National Perspective." New York: Routledge, pp 309–323.

Olweus D. 1989. The Olweus Bully/Victim Questionnaire. Bergen. Mimeograph.

Olweus D. 1991. Bully/victim problems among schoolchildren: Basic facts and effects of a school-based intervention program. In Pepler D, Rubin KH (eds): "The Development and Treatment of Childhood Aggression." Hillsadel, NJ: Erlbaum, pp 411–448.

Olweus D. 2001. Sweden. In Juvonen J, Graham S (eds): "Peer Harassment in School: The Plight of the Vulnerable and Victimized". New York: Guildford Press.

Oude Nijhuis M. 2001. Pesten als groepsprocess. Een onderzoek naar validering van de Participant Role Scales. [Bullying as a group process. A study into the validity of the Participant Role Scales] Unpublished master's thesis. Department of Education, Free University of Amsterdam.

Pellegrini AD. 1996. "Observing Children." Hilsdale, NJ: Erlbaum.

Pellegrini AD. 2001. Sampling instances of victimization in middle school: A methodological comparison. In Juvonen J, Graham S (eds): "Peer Harassment in School. The Plight of the Vulnerable and Victimized." New York: Guildford, pp 125–144.

Pellegrini AD. 2002. Affiliative and aggressive dimensions of dominance and possible functions during early adolescence. Aggress Violent Behav 7:21–31.

Pellegrini AD, Bartini M. 2000. An empirical comparison of methods of sampling aggression and victimization in school settings. J Educ Psychol 92:360–366.

Pepler DJ, Craig WM, Roberts WL. 1998. Observations of aggressive and nonaggressive children in the school playground. Merrill-Palmer Quart 44:55–76.

Perry DG, Kusel SJ, Perry LC. 1988. Victims of peer aggression. Dev Psychol 24:807–814.

Rigby K. 1996. "Bullying in Schools and What to do About it." Melbourne: ACER.

Salmivalli C, Lagerspetz K, Björkquist K, Österman K, Kaukiainen A. 1996. Bullying as a group process: Participant roles and their relations to social status within the group. Aggr Behav 22:1–15.

Salmivalli C, Lappalainen M, Lagerspetz K. 1998. Stability and change of behavior in connection with bullying in schools: A two-year follow-up. Aggr Behav 24:205–218.

Schwartz D, Proctor LJ. 2001. Community violence exposure and children's social adjustment in the school peer group: The mediating roles of emotion regulation and social cognition. J Consult Clin Psychol 68:670–683.

Schwartz D, Proctor LJ, Chien DH. 2001a. The aggressive victim of bullying: Emotional and behavioral dysregulation as a pathway to victimization by peers. In Juvonen J, Graham S (eds): "Peer Harassment in School. The Plight of the Vulnerable and Victimized." New York: Guildford, pp 147–174.

Schwartz D, Chang L, Farver JM. 2001b. Correlates of victimization in Chinese children's peer groups. Dev Psychol 37:520–532.

Slee PhT. 1995. Peer victimization and its relationship to depression among Australian primary school students. Pers Indiv Differ 18:57–62.

Smith PK, Sharp S (eds): 1994. "School Bullying: Insights and Perspectives." London: Routledge.

Smith PK, Madsen KC, Moody JC. 1999. What causes the age decline in reports of being bullied at school? Towards a developmental analysis of risks of being bullied. Educ Res 41:267–285.

Smith PK, Cowie H, Olafson RF, Liefooghe APD, with the collaboration of Ameida A, Araki H, del Barrio C, Costabile A, Dekleva B, Houndoumadi A, Kim K, Olafsson RP, Ortega R, Pain J, Pateraki L, Schäfer M, Singer M, Smorti A, Toda Y, Tomasson H, Wenxin Z. 2002. Definitions of bullying: A comparison of terms used, and age and gender differences, in a fourteen-country international comparison. Child Dev 73:1119–1133.

Sutton J, Smith PK. 1999. Bullying as a group process: An adaptation of the participant role approach. Aggr Behav 25:97–111.

Sutton J, Smith PK, Swettenham J. 1999. Bullying and theory of mind: A critique of the social skills deficit view of anti-social behavior. Soc Dev 8:117–127.

Swets JA. 1996. "Signal Detection Theory and ROC Analysis in Psychology and Diagnostics (Collected Papers, 1996)." Mahwah, NJ: Erlbaum.

Whitney L, Smith PK. 1993. A survey of the nature and extent of bullying in junior/middle and secondary schools. Educ Res 35:3–25.

Wolke D, Woods S, Bloomfield L, Karstadt L. 2000. The association between direct and relational bullying and behavior problems among primary school children. J Child Psychol Psychiatry 41:989–1002.